ON THE ACCURACY OF 3D LANDSCAPES FROM UAV IMAGE DATA

Koen Douterloigne[*]*, Sidharta Gautama, Wilfried Philips*

Department of Telecommunications and Information Processing (UGent-TELIN-IPI-IBBT)
Ghent University, St-Pietersnieuwstraat 41, B-9000 Ghent, Belgium
Koen.Douterloigne@telin.ugent.be

ABSTRACT

Our surroundings change all the time. Applications that require 3D models of a changing terrain, such as urban planning, are becoming ever more demanding with respect to the cost to create them and the accuracy of the result. A novel, cheap and fast solution for this problem is given by a UAV to take aerial images of the terrain in question, in combination with structure from motion algorithms to create a 3D model from those aerial images. However the question remains whether these on-the-fly 3D maps can match the accuracy of classical surveyor based models, which require more time to create. In this paper we investigate this question, and find that under certain conditions the accuracy of the UAV based model matches the accuracy of surveyor generated measurements.

Index Terms— 3D reconstruction, UAV, accuracy, evaluation

1. INTRODUCTION

Aerial images taken by an unmanned aerial vehicle (UAV) can be used for many purposes. The most obvious ones are, just like aerial images from other platforms, the generation of large orthophotos and the surveillance of ground targets. In addition, structure from motion algorithms [1] have enabled the creation of dense digital terrain models [2]. This gives us a complete three dimensional (3D) model of the overflown terrain. As real world applications using 3D models become more demanding, a rough approximation is not good enough anymore. For example, planning urban environments and infrastructures requires knowledge of the terrain up to submeter accuracy, a task which is currently performed by surveyors. We can also obtain this information by taking images with a UAV, which offers the benefit of being both cheaper and faster than surveyors. However in order to position these images, the position of the UAV must be known exactly at the moment each picture was taken. Given that due to weight constraints the GPS carried onboard the UAV has limited accuracy, and that there can be a time delay between estimating the position with GPS and actually taking the picture, the question is how accurate we can know this position. More specifically, is the obtained accuracy high enough to complement (or even replace) surveyors as the method of choice for applications requiring high precision 3D models.

Previously, work has been performed comparing accuracy and completeness of dense 3D reconstructions [3]. However, this work is limited to the final step of the workflow shown in figure 1, and also does not consider the possibility of adding prior knowledge to improve the reconstruction. In this paper we evaluate the 3D reconstruction by comparing a model generated from 3D coordinates measured by a surveyor, to the aforementioned structure from motion from UAV images, increased by the UAV's internal GPS. We also investigate the effect of adding manually measured ground control points. Several other methods to obtain a 3D model exist, among others time-of-flight cameras, structured light, or laser measurements. However these are either low-resolution, or impossible to mount on a UAV, and for that reason we did not consider them in our comparison.

The rest of this paper is arranged as follows. First we go over the methods used to go from a set of images plus GPS to a georeferenced 3D model. Next we describe the setup we used to evaluate the accuracy of a real world application. We then discuss some results, and end with a conclusion.

2. METHODOLOGY

The workflow to reconstruct a 3D model from a set of images is shown in figure 1. First distinctive feature points are extracted from all images, along with a feature descriptor vector which collects statistics of a window around the feature. A wide range of features exist, among which SIFT [4] and SURF [5] are well known. We then try to match feature points corresponding to the same physical object in as much images as possible. This is done by computing the Euclidean distance between the feature descriptors. Two points are said to match when their descriptors are close together in n-dimensional Eu-

[*]This work was performed as part of an IWT project between Ghent University and Gatewing.

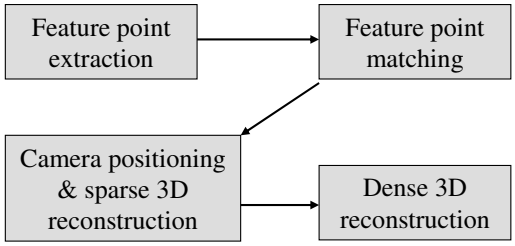


Fig. 1. Schematic representation of the steps required to go from a set of images to a dense 3D model.

clidean space, where n is the size of the descriptor vector. The definition of closeness is taken according to the method described in [4], where the distance d_1 between a point and its closest neigbour is compared to d_2, the distance between that point and its second closest neighbour. A match is retained when $\frac{d_1}{d_2} < 0.6$.

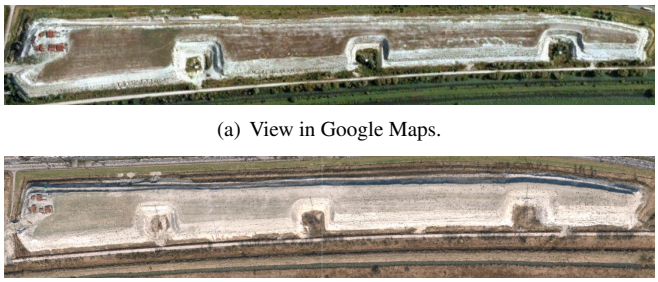
The number of computations required to match feature points between all images is quadratic in both the amount of images and the amount of points, and can thus take a lot of time for large scenes. In [1] the speed is improved by using approximate nearest neighbour searching [6]. When dealing with images from a UAV, we can also use the GPS to roughly position the images in space, thus limiting the set of possible matches. Practically, two points will not be matched when the GPS position of the images in which they were found are too far apart.

Once we know a large set of corresponding points, we can start the 3D reconstruction. In projective coordinates, a 3D point $\mathbf{X}(x, y, z, w)$ is projected onto a 2D image pixel $\mathbf{x}(x, y, w)$ under the following formula:

$$\lambda \mathbf{x} = \mathbf{M}\mathbf{X}, \tag{1}$$

where λ is a scale factor, and $\mathbf{M}$ is the *camera matrix*, an arbitrary homogeneous 3 x 4 matrix with rank 3, depending on 11 parameters [7]. This matrix contains information about the camera position and orientation (equivalent to the position and orientation of the UAV at the time of taking the picture), as well as some parameters describing the camera's optical properties, such as its focal length, its principal point, and its aspect ratio. Together these parameters uniquely determine the scene visible at a certain time. Note however that formula (1) does not take the optical aberrations or lens distortions of the camera into account. In order to keep the explanation brief, it is assumed that any lens distortion has been removed in advance, using e.g. the technique described in [8], based on the work of [9].

The 3D reconstruction then comes down to finding values for all camera matrices $\mathbf{M}_i$ for $i = 1..m$ pictures, and all points $\mathbf{X}_j$ for $j = 1..n$. We will further write the projection of point $\mathbf{X}_j$ onto image $\mathbf{M}_i$ as $\mathbf{x}_{i,j}$. The constraints are then given by the requirement that the distance between the



(a) View in Google Maps.

(b) Top-down view of the reconstructed 3D model.

Fig. 2. Orthophotos of the area used for evaluating the accuracy of the 3D reconstruction.

position of a feature point $\hat{\mathbf{x}}_{i,j}$ determined from the feature extraction, and its calculated position $\mathbf{x}_{i,j} = \mathbf{M}_i\mathbf{X}_j$, should be as small as possible. With $d(\mathbf{x}, \mathbf{y})$ denoting the Euclidean distance between 2D points $\mathbf{x}$ and $\mathbf{y}$, we can rewrite this as:

$$\min_{\mathbf{M}_i, \mathbf{X}_j} \sum_{i=1}^{m} \sum_{j=1}^{n} d\left(\hat{\mathbf{x}}_{i,j} - \mathbf{M}_i\mathbf{X}_j\right)^2 \tag{2}$$

Solving this nonlinear equation is not a simple task given the large number of variables involved. Even a simple scene quickly has thousands of 3D points. For this reason bundle adjustment is used [10], which solves (2) by exploiting the sparsity in the equations, which stems from the fact that all $\mathbf{X}_j$ do not influence eachother. The same goes for all $\mathbf{M}_i$. The formula is then optimized with the Levenberg-Marquardt algorithm [11]. For this algorithm to converge it is very important that we start from an approximate solution for both the camera positions and the coordinates of the points. We combine the standard RANSAC based approach of approximately positioning the images with respect to eachother, with the absolute (albeit inaccurate) position information obtained from the GPS of the UAV.

In a final step we use the solution of (2) as the input to a dense reconstruction, based on multi-view stereopsis [2, 7]. It is on this dense reconstruction that we will run our evaluation. Note that the accuracy of the result is more than simply the performance of the last step in the process, which has been thoroughly evaluated in [3].

3. TEST SETUP

For the evaluation of the accuracy of the reconstructed 3D model we limit ourselves to a specific site, namely an area containing a long, flat-topped and man-made hill, measuring about 1500 x 300 m. A satellite picture of the area as seen in Google Maps is shown in figure 2. Furthermore, 15 yellow cross shaped markers were added on the site and measured very precisely using differential GPS and the Flemish FLEPOS post-processing system [12], giving their position up to 10 cm. Even though there is still an error on their measured

position, we use these markers as a ground truth in the comparisons.

With the markers in place, a UAV from the company Gatewing flew over this terrain, taking a total of 439 images in 5 flight lines with a 90% overlap in a flight line and a 60% overlap between flight lines, allowing for good image matching and good stereovision. The UAV flew at an average altitude of 150 m and took 10 megapixel pictures, resulting in an average pixel size of about 5 cm. This ensured that the markers were well visible in the pictures, and that every marker was visible in at least 5 images. Next, the methods described in section 2 were applied. The exact center of the markers visible in the images were determined manually, and then also taken along in the bundle adjustment, giving us a computed 3D coordinate for each marker. Comparing this computed result with the measured position gives us a quantitative indication of the accuracy of the computations.

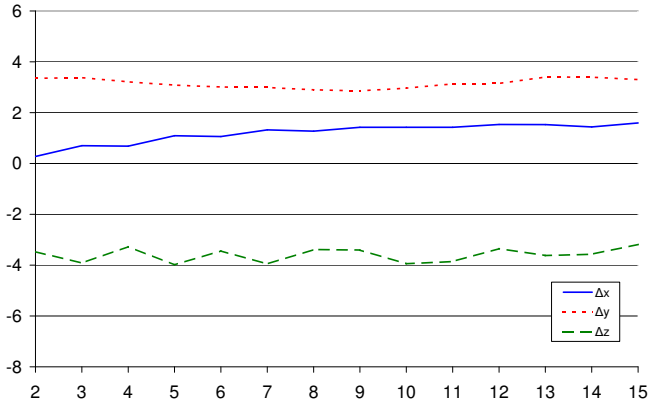
4. RESULTS

In figure 3(a) the differences in meter between measured and computed marker positions are shown, split into Δx, Δy and Δz. Marker number 1 was not used because it was not visible in enough images. We see that there is a quite large deviation of up to 4 meters from the ground truth. This is explained by the error on the GPS measurements in the UAV, as well as the time delay between a GPS snapshot and capturing an image. Unfortunately we see no way to solve this without the use of extra information. The position of the UAV is, under these conditions, our only link to a georeferenced model, and any error on this position will inevitably propagate to the 3D model.

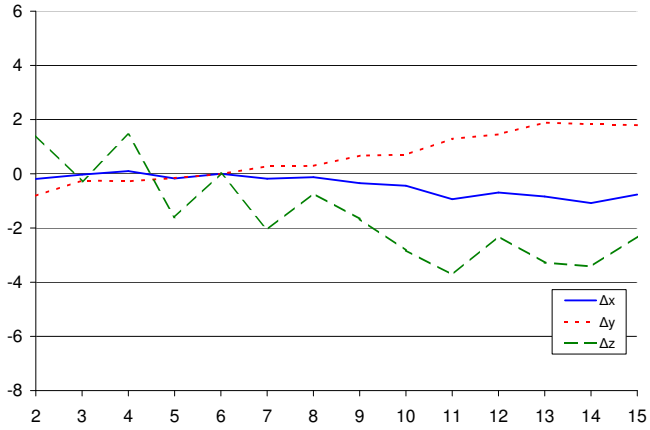
When we use some ground truth information in the form of the computed marker position, the results improve markedly. This is shown in figures 3(b) and 3(c). Adding one such marker or ground control point (GCP) pulls the entire model more to the correct location near that marker. Obviously the marker itself will have a perfect position. Parts of the model that lay far away however still use the initial, GPS based position of the images, and retain an error of several meters.

Adding more GCPs further improves the result. It turns out that adding 4 of the 15 points is sufficient to negate the effect of the biased GPS. The errors on the marker positions are now in the range of 10 to 20 cm, which is close to the accuracy achieved by the diffential GPS measurement.

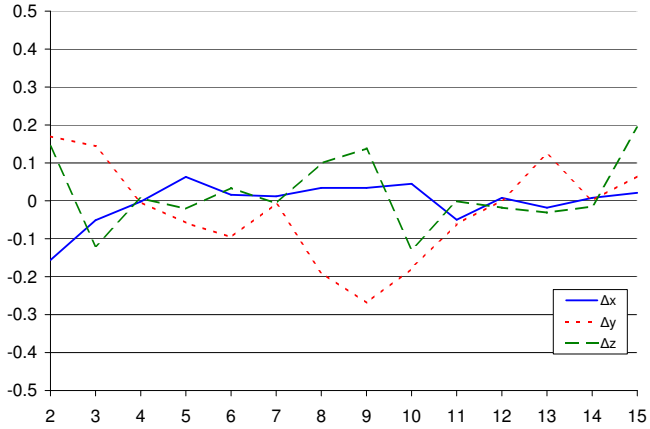
Finally, in figure 4 we show part of the densely reconstructed 3D model after meshing with Delaunay triangulation. Some noise is visible, especially on the road where it is harder to find corresponding pixels due to a lack of details. This noise also causes the fluctuations that are visible in figure 3(c). Future work on this topic may improve the results further, through smart noise correction or better pixel matching.



(a) No GCPs added



(b) GCP 6 added



(c) GCPs 4, 7, 12 and 14 added

Fig. 3. The difference in meter between measured and computed marker positions, showing the effect of adding ground control points (GCPs) into the bundle adjustment.

5. CONCLUSION

In this paper we have evaluated the state of the art 3D reconstruction methods applied to UAV images positioned using

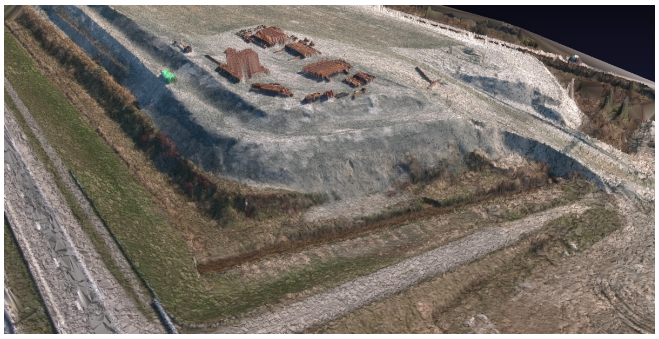


Fig. 4. Close-up of the densely generated 3D model, showing the left side of the hill from figure 2. This visible subsection contains about 1 million vertices.

GPS information and surveyed ground control points. It was found that the 3D model has an average accuracy of 10 to 20 cm in all directions, for a pixel size of 5 cm. We must note however that this result is obtained with the inclusion of a few ground control points, spread evenly over the terrain. This implies that ground based surveying is still required, but only at a fraction of the time required without a UAV, as only a fraction of the points must be surveyed. When no ground control points are used, the accuracy is governed by the accuracy of the GPS of the UAV, and is about 5m.

The accuracy of 10 to 20 cm is sufficient for many practical applications, however there is still room for improvement. For example, adding more ground control points will further increase the precision, up to maximally the precision of the differential GPS measurement. This of course has to be balanced by the amount of manual labor required, both to place and measure the markers, and to add them to the workflow. Additionally, even more advanced pixel matching methods can also further improve the results, at the cost of computation time.

6. ACKNOWLEDGEMENTS

We would like to thank the people at Gatewing for the use of their datasets, and for valuable discussions about the accuracy of the obtained reconstruction results.

7. REFERENCES

[1] M.I. A. Lourakis and A.A. Argyros, "SBA: A Software Package for Generic Sparse Bundle Adjustment," *ACM Trans. Math. Software*, vol. 36, no. 1, pp. 1–30, 2009.

[2] Yasutaka Furukawa and Jean Ponce, "Accurate, Dense, and Robust Multi-View Stereopsis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009.

[3] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2006, pp. 519–528, IEEE Computer Society.

[4] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[6] Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, 1998.

[7] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2003.

[8] K. Douterloigne, S. Gautama, and W. Philips, "Fully automatic and robust UAV camera calibration using chessboard patterns," in *IEEE International Geoscience and Remote Sensing Symposium*, July 2009, pp. 551–554.

[9] Zhengyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[10] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon, "Bundle adjustment - a modern synthesis," in *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, London, UK, 2000, pp. 298–372, Springer-Verlag.

[11] Donald W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[12] Björn De Vidts and Bart Dierickx, *Performing GPS measurements with Flemish Positioning Service (FLEPOS)*, AGIV, 2008.